Matérn-based nonstationary cross-covariance models for global processes

Mikyoung Jun
Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, USA

Abstract

Many spatial processes in environmental applications, such as climate variables and climate model errors on a global scale, exhibit complex nonstationary dependence structure, not only in their marginal covariance but also in their cross-covariance. Flexible cross-covariance models for processes on a global scale are critical for an accurate description of each spatial process as well as the cross-dependences between them and also for improved predictions. We propose various ways to produce cross-covariance models, based on the Matérn covariance model class, that are suitable for describing prominent nonstationary characteristics of the global processes. In particular, we seek nonstationary versions of Matérn covariance models whose smoothness parameters vary over space, coupled with a differential operators approach for modeling large-scale nonstationarity. We compare their performance to the performance of some existing models in terms of the AIC and spatial predictions in two applications: joint modeling of surface temperature and precipitation, and joint modeling of errors in climate model ensembles.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

In environmental applications, much attention has recently been paid to the problem of modeling the cross-covariance of multivariate spatial and spatio-temporal processes and co-kriging (e.g. [21,14,7,17]). Geophysical applications often have multiple variables of interest that cover large areas of the earth. Examples include joint modeling of surface temperature and precipitation variables in climate problems [32,13] and joint modeling of multiple climate model outputs on a global scale [28]. Sang et al. [28] built a joint statistical model for five climate model outputs using a parametric cross-covariance function based on the linear model of coregionalization [34,9].

Significant effort has been devoted to developing valid cross-covariance functions; these include [1,11,13,25,2,19,23]. Gneiting et al. [11] proposed a multivariate version of Matérn covariance functions that were further extended by Apanasovich et al. [2]. Similarly to the Matérn covariance function for univariate processes, these covariance models are isotropic, but flexible, given only a small number of parameters. Although these covariance functions were originally developed for processes on a planar domain, that is, $\mathbb{R}^2$, they can be applied to processes on the surface of a sphere.

For a Matérn class to be a valid covariance function for processes on a sphere, the distance metric that is used is important. As reported in [10] (and the references therein), if we use the great circle distance, which is a natural distance metric on the surface of the sphere, then the Matérn class is valid if and only if $0 < \nu \leq 0.5$ for the smoothness parameter $\nu$. On the other hand, if we use chordal distance, the Matérn covariance function with $\nu > 0$ is valid on the surface of the sphere. Although chordal distance on the surface of the sphere may be less desirable due to its distortion of the distance, especially...
when two points are relatively far apart, we may nevertheless rely on chordal distance since many geophysical processes are too smooth to be modeled with a Matérn class with $\nu \leq 0.5$. Further, the distortion of the chordal distance in modeling covariance may be of less concern when the spatial dependence range is not too large. Recently, Du et al. [5] presented valid multivariate variogram models using the great circle distance for processes on a sphere that are also isotropic.

The isotropic Matérn covariance models in [11] and [2], as well as the models in [5], are clearly limited in modeling multivariate processes on a sphere in environmental problems since such processes often exhibit strong nonstationarity. In particular, those covariance models give constant (so-called) co-located correlation coefficients across the entire domain. Jun [13] showed that for global surface temperature and precipitation data, the estimated co-located correlation coefficient of the bivariate parsimonious Matérn model is close to zero when there is a clear (spatially varying) dependence between the two variables. Kleiber and Nychka [19] proposed a nonstationary version of the multivariate Matérn model that can produce flexible spatially varying parameters of the Matérn class, although its implied co-located correlation coefficient has a limited spatially varying structure. Along these lines, Kleiber and Genton [18] suggested a nonparametric method of post-processing the cross-correlation to allow the cross-correlation matrix to vary over space, and this method can be applied to parametric cross-covariance models, including the multivariate Matérn model. However, the method requires nugget effects in the cross-covariance model. Furthermore, the estimators for the spatially varying co-located cross-correlation coefficients are based on kernel-smoothed products of two processes at the same observation locations. Therefore, if the multivariate processes are not observed at the same locations, the observations cannot be used in the estimation of co-located cross-correlations.

In this paper, we explore various parametric approaches for nonstationary versions of Matérn models that do not require post-processing of cross-correlations and we compare them in two applications, focusing on the processes on a sphere. Regarding the nonstationary Matérn model discussed in [19], we demonstrate that allowing smoothness parameters to vary over space may be the key for a better fit, compared to other types of Matérn class. We also study some characteristics of the differential operators approach for multivariate processes that was studied in [13], and we propose an enhanced approach that gives nonstationary Matérn covariance with the parametric form of flexible spatially varying co-located cross-correlation coefficients. In the applications, we consider situation that include locations where not all processes are observed.

The rest of the paper is organized as follows. Section 2 discusses some of the existing Matérn-type cross-covariance models in greater detail. In Section 3, we study a nonstationary Matérn cross-covariance model with spatially varying smoothness. In Section 4, we present new model classes that couple nonstationary Matérn cross-covariance models with the differential operators approach. In Section 5, we apply our approach to two application problems involving joint modeling of two climate variables and the errors of climate model ensembles. We conclude the paper with discussions in Section 6.

2. Background

In this section, we study properties and limitations of existing Matérn-based cross-covariance models in the literature, particularly for the processes on a sphere.

2.1. The isotropic multivariate Matérn covariance model

Gneiting et al. [11] applied their multivariate Matérn model to the joint modeling of temperature and pressure observations (forecast error field) at 157 locations in the North American Pacific Northwest. Assuming that the means of both fields are zero, they fitted various (isotropic) versions of the multivariate Matérn model and compared them with the fitted results of the linear model of coregionalization. They showed that most of the cross-covariance models they considered do not make much difference in terms of the maximum likelihood fit or prediction accuracy, while they outperform the independent Matérn model due to the clear cross dependence of the two variables. However, in their comparison between the fitted covariances (both marginal and cross-covariance) and the empirical ones, they showed that there is a big discrepancy between the fitted values and the empirical values, especially near the origin. They explained that this type of disagreement is likely due to dependencies in the empirical covariance function and biases.

Although such a discrepancy may be partially explained by the dependence of empirical covariance, it may be due to a lack of flexibility in the covariance model. In Section 3.1, we show that we can significantly improve on this misfit by letting the smoothness of the process be a function of latitude and longitude.

2.2. The nonstationary multivariate Matérn covariance model

Stein [29] proposed a way to allow the smoothness parameters and geometric anisotropy of the Matérn covariance function for univariate processes to vary over space (see Theorem 1 of [24] for a more general case). This idea has recently been applied to the multivariate case by Kleiber and Nychka [19], who show that the same approach results in a valid multivariate Matérn covariance function class with this nonstationary property. However, their model for the cross-correlation coefficient uses a kernel-smoothed estimate based on a pairwise product of multivariate processes, similarly to [18]. Therefore, if there are locations where not all the processes are observed, the model may not be able to use the available observations over these locations.
In their application to joint modeling of temperature and precipitation data from regional climate models, Kleiber and Nychka [19] used fixed smoothness parameters over space. Although these constant smoothness parameters for the regional climate model output may not be restrictive, they may be too restrictive for global climate variables. In Section 5, we demonstrate that allowing smoothness parameters to be different over land and sea or to depend on latitude tends to improve the fit of the models significantly.

2.3. The differential operators approach for multivariate processes on a sphere

The models discussed in Sections 2.1 and 2.2 were not originally designed for processes on a sphere. Jun [13] proposed an approach for nonstationary cross-covariance models, which is especially suitable for processes on a global scale. The proposed model was applied to the joint modeling of surface temperature and precipitation data from a global climate model and it was shown that the model’s fit is overall significantly better in terms of the Akaike Information Criterion (AIC) and prediction than the isotropic bivariate Matérn model and the linear model of coregionalization.

Suppose that a multivariate process $Z(s) = \{Z_1(s), \ldots, Z_n(s)\}$ is observed over the globe (that is, $s \in \mathcal{S}^2$) and that each process $Z_i$ is observed at $\mathcal{L}_i = \mathcal{L} \cup \mathcal{L}^{(i)}$, where $\mathcal{L} = \{s_1, \ldots, s_n\}$, $\mathcal{L}^{(i)} = \{s_i^1, \ldots, s_i^{n_i}\}$, and $n_i, i = 1, \ldots, N$, is a non-negative integer. Here $\mathcal{L}$ is the maximum set of locations where all the processes $Z_i$ are observed, and $\mathcal{L}^{(i)}$ is the set of locations not in $\mathcal{L}$ where $Z_i$ is observed. For some $i$, the set $\mathcal{L}^{(i)}$ may be empty, and for some $i$ and $j$ ($i \neq j$), $\mathcal{L}^{(i)}$ and $\mathcal{L}^{(j)}$ may have a non-empty intersection. When all the variables are observed in the same spatial locations, $\mathcal{L}^{(i)}$ is empty for all $i$.

Let $s = (L, l) \in \mathcal{S}^2$, where $L$ is the latitude and $l$ is the longitude, and let $Z^s$ be a process on the sphere with an isotropic Matérn covariance structure, through the restriction of the Matérn covariance function in $\mathbb{R}^3 \times \mathbb{R}^3$ onto $\mathcal{S}^2 \times \mathcal{S}^2$ (see (7) of [10] for more details). Then Jun [13] wrote

$$Z_i(L, l) = \{A_i(L) \partial / \partial L + B_i(L) \partial / \partial l\} Z^s(L, l), \quad i = 1, \ldots, N. \quad (1)$$

Here, the partial derivatives are defined in the $L$-2 sense, and thus we need $\nu > 1$ for the smoothness parameter $\nu$ of the Matérn covariance of $Z^s$ in order for (1) to be defined. The coefficients $A_i$ and $B_i$ are deterministic functions depending on the latitude, and in particular, [13] used finite linear combinations of Legendre polynomials. That is, if we let

$$P(p_0, \ldots, p_m; \lambda) = \sum_{j=0}^{m} p_j P_j(\sin L),$$

where $P_j$ denotes the Legendre polynomial of order $j$, and $p_j \in \mathbb{R}$ ($j = 0, \ldots, m$), then Jun [13] used

$$A_i(L) = P(a_{i0}, \ldots, a_{im}; \lambda) \quad \text{and} \quad B_i(L) = P(b_{i0}, \ldots, b_{in}; \lambda),$$

for $m, n \in \mathbb{N}$. The constants, $a_{ij}, b_{ij} (\in \mathbb{R})$, are estimated along with the other covariance parameters.

One problem with the model in (1) is that it does not respect the curvature of the sphere. That is, the differential operators are effectively applied over a cylinder rather than a ball and this distorts the spherical domain, especially near the poles. This problem possibly caused the poor fit of longitudinal irreversibility discussed in [13]. Another problem with the model in (1) is that, as discussed in [15,16], it is generally not mean-square continuous at both poles. Furthermore, we shall show in Section 4.2 that the co-located cross-correlation under the model in (1) with (3) at both poles has magnitude 1, which is unrealistic for most real applications.

3. Models with spatially varying smoothness

The first model that we discuss here is a straightforward application of the model in [19] for processes on a sphere, but we focus on spatially varying the smoothness parameter, along with varying the marginal variance over latitude. That is, if we let $s_k = (L_k, l_k) \in \mathcal{S}^2$ and $K_{ij}(s_1, s_2) = \text{cov}[Z_i(s_1), Z_j(s_2)]$, we write

$$K_{ij}(s_1, s_2) = \sigma_i(L_1) \sigma_j(L_2) M_{ij}(s_1, s_2) (D(s_1, s_2) / \xi_i),$$
$$K_{ij}(s_1, s_2) = \beta_{ij} \sigma_i(L_1) \sigma_j(L_2) M_{ij}(s_1, s_2) (D(s_1, s_2) / \xi_j), \quad i \neq j. \quad (4)$$

for $i, j = 1, \ldots, N$, where $\sigma_i : [-\pi/2, \pi/2] \rightarrow [0, \infty)$ is a function of latitude, $M_{ij}(x) = x^\nu K_{ij}(x)$, with $K_{ij}$ a modified Bessel function of the second kind with order $\nu$, $v_{ij}(s_1, s_2) = \nu_i(s_1) + \nu_j(s_2)/2$, $v_{ij}(s_1, s_2) = \{v_i(s_1, s_2) + v_j(s_1, s_2)\}/2$, and $D$ is the chordal distance between the two points on the sphere. The parameters $\xi_i$ and $\xi_j$ are range parameters, and $\xi_{ij} = (\xi_i + \xi_j)/2$ for all $i$ and $j$. The function $v_{ij}(\cdot)$ is a positive function that varies over space. We also need an $N \times N$ matrix, where the $(i,j)$th entry $\beta_{ij}$ is such that $\beta_{ii} = 1$ and $-1 \leq \beta_{ij} \leq 1$, to be symmetric and nonnegative definite.

Allowing smoothness to vary over space may be effective when we focus on fitting the marginal and cross-covariance near the origin [30], and therefore this approach can potentially be useful for the fits of the marginal and cross-covariance near the origin (see Fig. 1). We consider two versions of $v_{ij}$ in this paper; one is expressed as a linear combination of Legendre polynomials depending on the latitude, and the other takes different values over the land and the sea. In any case, we need the resulting smoothness parameter values to always be positive in order to keep the covariance model valid.
The model in (4) gives the co-located cross-correlation in the following form [19]:
\[
\text{cor}(Z_i(s), Z_j(s)) = \beta_{ij} \Gamma\left(\nu_i(s) + \nu_j(s)\right) / 2 \Gamma^{-1/2}(\nu_i(s)) \Gamma^{-1/2}(\nu_j(s)).
\] (5)

Note this is not necessarily a constant over space.

3.1. Example from [11]

We fitted simple versions of the model in (4) to the bivariate data used in [11] and compared our fitted results to those from the isotropic version of the model presented in [11]. In particular, we let \( \sigma_i \) be a constant and \( \nu_i \) a linear function of longitude and latitude. We let \( \xi_i = \xi \) for all \( i \), so that we could check the effectiveness of spatially varying the smoothness parameters. The resulting variances and covariance as well as the cross-correlation of the bivariate process depend on the longitude and latitude, although they have a limited structure (see, for example, (5)). As a comparison, we also fitted a model version where \( \sigma_i \) in (4) is a linear function of longitude and latitude (while \( \nu_i \) is kept constant).

Similarly to [11], these cross-covariance models are fitted using the maximum likelihood estimation method that assumes that both variables follow a joint Gaussian distribution and have the mean zero. The parsimonious Matérn model (call it ISO) has 8 parameters (including nugget effects for both variables), a nonstationary version with spatially varying smoothness (call it VS) has 12 parameters, and the other nonstationary version with spatially varying covariance and constant smoothness parameters (call it VV) has 12 parameters. The two nonstationary versions (VS and VV) increased the maximum loglikelihood values by 24.61 and 12.66, respectively, compared to the bivariate parsimonious Matérn model (ISO). In terms of the \( \text{aic} \), the model with spatially varying smoothness (VS) fits the data best among the three models considered.

Fig. 1 shows the comparison between the empirical covariances and fitted values from the three covariance models. For the fitted values of the two nonstationary models, we summarize the fitted covariances as boxplots at given distance lags (whisks are not displayed). Fig. 1 clearly shows that the fit to the marginal covariance as well as the cross-covariance is significantly improved with the nonstationary model with varying smoothness (VS). Large discrepancies near the origin shown in [11] (ISO) are significantly reduced with the nonstationary model VS. On the other hand, the nonstationary model VV produces much worse fit, even compared to the fit of the isotropic model, for pressure variable. This is a clear sign that spatially varying smoothness is the key to fitting the covariance near the origin for this data set.

Fig. 2 shows the fitted values for the smoothness parameters for the nonstationary model with spatially varying smoothness. The smoothness parameter values for the pressure variable are clearly a linear function of the latitude but they seem independent of longitude. On the other hand, the smoothness parameter values for the temperature variable show a linear relationship with both latitude and longitude. The linear pattern of the smoothness parameter against longitude for the temperature variable may suggest that the field is smoother over the sea than over the land (since the left half of the domain is mostly sea and the right half is mostly land).

In the comparison of VV and VS above, VS was found to be more effective in terms of model fitting. However, this does not imply that spatially varying smoothness is enough for most observations. Many environmental data exhibit obvious spatially varying variances (see [15,16,13,28]), and in general we need to account for such characteristics in addition to spatially varying smoothness. Most of the nonstationary covariance models considered in Section 5 have both spatially varying variance and spatially varying smoothness components.

4. Coupling the nonstationary Matérn model with the differential operators approach

Another limitation of the model in (4) is that its correlation cannot change signs (since the univariate Matérn covariance function only yields positive spatial correlations). In particular, the co-located cross-correlation in (5) is always either non-
Fig. 2. Fitted values of the smoothness parameters from the nonstationary cross-covariance model with spatially varying smoothness: (a) fitted smoothness parameter values for pressure against latitude, where the gray scale gives the longitude values (degrees); (b) same as (a), except the fitted values are shown against longitude and the gray scale gives the latitude values (degrees); (d) and (e) are the same as (a) and (b), respectively, for the temperature variable; (c) and (f) show the fitted smoothness values over the spatial domain for pressure and temperature, respectively.

negative or non-positive. Furthermore, apart from the $\sigma_i$s, the model in (4) gives a symmetric cross-covariance structure. In particular, the cross-correlation structure is symmetric. We say that a multivariate process $Z$ is asymmetric if

$$\text{cov}[Z_i(L_1, l_1), Z_j(L_2, l_2)] \neq \text{cov}[Z_i(L_2, l_2), Z_j(L_1, l_1)],$$

for some $i \neq j, l_1, l_2, l_1, l_2$, and longitudinally irreversible if

$$\text{cov}[Z_i(L, l_1), Z_j(L, l_2)] \neq \text{cov}[Z_i(L, l_2), Z_j(L, l_1)],$$

for some $i \neq j, l_1, l_2$. Although the covariance matrix for a univariate process ought to be symmetric, it has been shown that many geophysical processes produce asymmetric cross-covariance structures (e.g. [20]).

On the other hand, regarding the model with the differential operators approach given in (1) that is not necessarily symmetric, it can be seen from the following proposition and its corollary that it not only produces mean-square discontinuity at the poles, but also an unrealistic co-located cross-correlation at the poles.

**Proposition 1.** If for a bivariate Gaussian process, $Z(L, l) = \{Z_1(L, l), Z_2(L, l)\}$ on the sphere, each $Z_i$ is modeled as in (1), then the co-located cross-correlation of $Z_1$ and $Z_2$ is given by

$$\text{cor}[Z_1(L, l), Z_2(L, l)] = \{-A_1(L)A_2(L) - B_1(L)B_2(L) \cos^2 L\}[A_1^2(L) + B_1^2(L) \cos^2 L]^{-1/2}\{A_2^2(L) + B_2^2(L) \cos^2 L\}^{-1/2}.$$

See the appendix for the proof of the proposition.
Corollary 1. For a bivariate process modeled as in (1) with (3),

\[ \lim_{l \to \pm \pi / 2} |\text{cor} [Z_1(l, 1), Z_2(l, b)]| = 1. \]

4.1. The coupled model

We now propose to apply the differential operators approach studied in [15, 16, 13] to the nonstationary Matérn model in (4), specifically with spatially varying smoothness parameters, to achieve cross-covariance models that allow asymmetry and a more flexible covariance structure. The cross-correlations are parametrically modeled and thus do not require post-processing of the model for a nonstationary structure. Even if there are locations where not all processes are observed, we may take full advantage of all the observations through the parametric cross-covariance model. Unlike the models discussed earlier, the cross-correlations can change signs in some parts of the spatial domain.

We combine the differential operators approach used in [13] with the multivariate Matérn class in (4). That is, we now consider

\[ Z_i(l, i) = \{ A_i(l) \partial / \partial L + B_i(L) \partial / \partial l \} Z^o_i(l, L), \]  

where \( Z^o_i(l, L), i = 1, \ldots, N \) is a multivariate process with the covariance structure given in (4). With (7), we achieve a less restrictive cross-covariance structure than (1), and in particular, the co-located cross-correlation is not necessarily 1, as shown in the following corollary.

Corollary 2. For a bivariate process modeled as in (7) with (3),

\[ \lim_{l \to \pm \pi / 2} |\text{cor} [Z_1(l, 1), Z_2(l, b)]| = \beta_{12}. \]

4.2. Choice of \( A_i \) and \( B_i \)

To overcome the limitation of the choice in (3) regarding the curvature of the earth as discussed in Section 2, we consider

\[ A_i(l) = P(a_{i,0}, \ldots, a_{i,m}; L), \quad B_i(L) = P(b_{i,0}, \ldots, b_{i,n}; L) / \cos L. \]  

The cosine term in \( B_i \) is necessary to respect the curvature of the earth.

On the other hand, an issue for the models in (1) or (7) (with (3) or (8)) is their discontinuity at the poles. For example, from the result in [15], we need

\[ \lim_{l \to \pm \pi / 2} |A_i^2(L) + B_i^2(L) \cos^2 L| = 0, \]  

\[ \text{to have each } Z_i \text{ mean-square continuous at the poles. It is easy to see that the model in (7) with (8) is mean-square continuous at the poles if} \]

\[ \sum_{j=0}^{m} (-1)^j a_{i,j} = \sum_{j=0}^{n} (-1)^j b_{i,j} = 0, \quad i = 1, \ldots, N, \]

which is somewhat arbitrary. Instead, Hitczenko and Stein [12] have suggested using

\[ A_i(l) = \cos L P(a_{i,0}, \ldots, a_{i,m}; L), \quad B_i(l) = P(b_{i,0}, \ldots, b_{i,n}; L), \]  

so that the resulting process is mean-square continuous, with any \( a_{i,j}, b_{i,j} \in \mathbb{R} \). Note that similarly to (8), having a cosine term in \( A_i \) as in (10) also helps respect the curvature of the earth. The process as defined in (1) or (3) with the choices in (8) or (10) is well defined, since \( A_i(L) \) and \( B_i(L) \cos L \) are bounded for all \( L \in [-\pi / 2, \pi / 2] \) [15].

Corollary 3. If a bivariate process \( Z(l, i) = \{ Z_1(l, 1), Z_2(l, 1) \} \) on the sphere has each \( Z_i \) modeled as in (1) or (7) and if we let \( A_i \) and \( B_i \) be defined as in (8) or (10), we have

\[ \text{cor} [Z_1(l, 1), Z_2(l, 1)] = -\gamma_{12} (A_1^2(L) B_2(L) + \bar{B}_1(L) \bar{B}_2(L)) (\bar{A}_1^2(L) + \bar{B}_1^2(L))^{-1/2} (\bar{A}_2^2(L) + \bar{B}_2^2(L))^{-1/2}, \]

with \( \bar{A}_i(L) = P(a_{i,0}, \ldots, a_{i,m}; L), \bar{B}_i(L) = P(b_{i,0}, \ldots, b_{i,n}; L), \text{ and } a_{i,j}, b_{i,j} \in \mathbb{R} \). Furthermore, we have

\[ \lim_{l \to \pm \pi / 2} |\text{cor} [Z_1(l, 1), Z_2(l, 1)]| = -\gamma_{12} (\alpha_1 \alpha_2 + \delta_1 \delta_2) (\alpha_1^2 + \delta_1^2)^{-1/2} (\alpha_2^2 + \delta_2^2)^{-1/2}, \]

where \( \alpha_i = \sum_{j=0}^{m} (-1)^j a_{i,j} \) and \( \delta_i = \sum_{j=0}^{n} (-1)^j b_{i,j}, i = 1, 2. \) The parameter \( \gamma_{12} = 1 \) for model (1), and \( \gamma_{12} = \beta_{12} \) for model (7).

Although under model (7), the coefficients in the form of either (8) or (10) give the same co-located cross-correlations, (8) does not in general produce mean-square continuous processes. This seems to matter in real applications (see Section 5 for a comparison of the two choices (8) and (10) for \( A_i \) and \( B_i \)).
5. Applications

In the following two examples of applications, we compare the various covariance models developed above in the joint modeling of climate variables and in climate model errors. The covariance models can be categorized as in Table 1. We will further describe how some of the models are used in the two applications and, in particular, the identifiability conditions on some of the parameters. We do not include the nugget effects for the variables in either example.

5.1. Joint modeling of surface temperature and precipitation

The cross dependence of surface temperature and precipitation is of particular interest in climate impact studies and there has been some interest in studying the relationship between the two variables (e.g., [26, 33]) and building joint spatial models of the two variables (e.g., [32, 13]). We apply the various covariance models presented in Sections 3 and 4 to model climate model outputs for the two variables. Similarly to [13], we consider 5-month averages of the two climate variables for the Northern winter months (November through March), averaged from 1970 to 1999. We use the residuals of the two variables after separately fitting spatially varying mean structures for each variable. See [13] for more details on how the spatial mean structures are fitted for the two variables.

Due to the huge size of the original climate model outputs, we randomly sampled 1000 locations (grid pixels) common for the two variables (i.e., \( L \)) from the original grid, with the resolution 128 \( \times \) 256 for each variable. We sampled additional 100 distinct locations for each variable (i.e., \( L^{(i)} \)). When sampling these locations, if we randomly sample from the spatial domain of \((-180^\circ, 180^\circ) \times (-90^\circ, 90^\circ)\) (i.e. longitude \( \times \) latitude), it is possible to end up with clusters of points near the poles. To mitigate this problem, we ensured that only 10% of the sampled points were in the regions north of 60° N and south of 60° S (the area of these regions is roughly 10% of the surface of the earth). Fig. 3 shows the map of observations used for both variables. The maps over the sampled locations show similar patterns to the original maps over all of the locations [13].

Among the models in Table 1, we fitted the following models IM, M1, M2-2, M2-3, M2-4, M3-1, M3-2, M3-3, M3-4, and M4-1. For M2-2, we set \( p_{1,0} > 0 \) to avoid the identifiability problem. For M2-3, we need the transformation on \( v_i \) to ensure that the smoothness parameters are positive and not too large for numerical stability. For M2-4, we set \( \eta_{i,j} > 0 \) for all \( i, j \). This model allows each process to be smoother over the sea (when we fitted the data over land and sea separately, we always obtained larger smoothness parameter values over the sea). For M3, we set \( m = n = 2 \) for \( A_i \) and \( B_i, i = 1, 2 \). The parameters for \( Y_i \) are kept the same as those for \( X_i \), except for one parameter for the co-located cross-correlation coefficient. To avoid the identifiability problem, we set \( a_{i,0} > 0 \). For M4-1, we set \( m = n = 2 \). Further, we set var(\( Y_i \)) = var(\( W \)) = 1, \( r_{i,0} > 0 \) for \( i = 1, 2 \), and \( a_{1,0}, d_1 > 0 \), to avoid the identifiability problem. We also kept the spatial range parameters for the \( X_i, Y_i, S, \) and \( W \) the same. The smoothness parameters for the \( Y_i \) were kept the same for each \( i \).

We fitted the covariance models through maximum likelihood estimation with the assumption that the processes have a joint Gaussian distribution. The maximized loglikelihood values, the numbers of parameters, and AIC values are summarized in Table 2. Rather than the actual loglikelihood and AIC values, we present the increase in the loglikelihood values and the decrease in the AIC values, both compared to the independent, isotropic Matérn model (IM). The model that yields the biggest decrease in the AIC values (which is thus the most preferable one in terms of AIC) is M3-4. This may imply that for this data set, having different smoothness parameter values over land and sea, and having the marginal- and cross-covariance structure vary over latitude are useful. It is interesting to note, comparing the fits of M3-1 and M3-2, that the coefficients in the form (10) fit the data better than in the form (8). This may suggest that mean-square continuity in the model is important in fitting

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IM</td>
<td>Independent Matérn model</td>
</tr>
<tr>
<td>M1</td>
<td>Parsimonious bivariate Matérn model from [11]</td>
</tr>
<tr>
<td>M2</td>
<td>Bivariate model in (4) with</td>
</tr>
<tr>
<td>M2-1</td>
<td>( \sigma_i(s), \nu_i, \xi_1 = \xi_2 = \xi ) constants across space</td>
</tr>
<tr>
<td>M2-2</td>
<td>Same as M2-1, except (s = (L, l))</td>
</tr>
<tr>
<td>M2-3</td>
<td>( \sigma_i(s) =</td>
</tr>
<tr>
<td>M3</td>
<td>( Z_i(L, l) = X_i(L, l) + Y_i(L, l) ), and</td>
</tr>
<tr>
<td>M3-1</td>
<td>( X_i ) is modeled with (7) (( X_i^{\circ} ) is modeled with M1) and (8), spatial range for each ( X_i ) is kept the same, ( Y_i ) is modeled using M1, and ( X_i ) and ( Y_i ) are independent</td>
</tr>
<tr>
<td>M3-2</td>
<td>Same as M3-1, except we use (10) instead of (8)</td>
</tr>
<tr>
<td>M3-3</td>
<td>Same as M3-2, except ( X_i^{\circ} ) is modeled with M2-4</td>
</tr>
<tr>
<td>M3-4</td>
<td>Same as M3-3, except the spatial range parameters for each ( X_i^{\circ} ) are different</td>
</tr>
<tr>
<td>M4</td>
<td>( Z_i(L, l) = X_i(L, l) + P(r_{i,0}, r_{i,1}, r_{i,2}; L)Y_i(L, l) + d_iW(L, l) ) and</td>
</tr>
<tr>
<td>M4-1</td>
<td>( X_i ) is modeled with (1) and (10), ( Y_i ) and ( W ) are modeled with a univariate Matérn model, ( d_2 = 0 ), and ( X_5, Y_5, S, ) and ( W ) are independent</td>
</tr>
<tr>
<td>M4-2</td>
<td>Same as M4-1, except ( Y_i ) is modeled with M1</td>
</tr>
</tbody>
</table>
Fig. 3. Surface temperature and precipitation data. Each process is separately mean filtered through regression with spherical harmonics terms as in [13]. The figures on the left show the locations of $L$ and those on the right show the locations of $L^{(i)}$ ($i = 1$ for temperature and $i = 2$ for precipitation).

Table 2

Maximum loglikelihood values, numbers of parameters, and AIC values for covariance models fitted to surface temperature and precipitation data. The maximum loglikelihood values, number of parameters, and AIC values for IM, are $-1160.27$, $5$, and $-2332.06$, respectively. $\Delta$loglik gives the increase in the loglikelihood values and $\Delta$AIC gives the decrease in the AIC values, both compared to IM.

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>M2-2</th>
<th>M2-3</th>
<th>M2-4</th>
<th>M3-1</th>
<th>M3-2</th>
<th>M3-3</th>
<th>M3-4</th>
<th>M4-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta$loglik</td>
<td>3.14</td>
<td>369.40</td>
<td>117.77</td>
<td>112.71</td>
<td>340.22</td>
<td>372.24</td>
<td>542.24</td>
<td>552.70</td>
<td>395.13</td>
</tr>
<tr>
<td>Parameters</td>
<td>6</td>
<td>12</td>
<td>10</td>
<td>8</td>
<td>19</td>
<td>19</td>
<td>21</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td>$\Delta$AIC</td>
<td>4.28</td>
<td>724.80</td>
<td>225.54</td>
<td>219.42</td>
<td>652.44</td>
<td>716.48</td>
<td>1052.48</td>
<td>1071.4</td>
<td>754.26</td>
</tr>
</tbody>
</table>

Table 3 gives fitted smoothness parameter values (along with their asymptotic standard errors) over the land and the sea for various covariance models. The asymptotic standard errors are obtained from the square roots of the diagonal elements of the inverse Hessian matrices, evaluated at the maximum likelihood estimates. Interestingly, IM and M1 produce nearly identical smoothness parameter values, while M2-2 does not. The estimated smoothness parameter values over land and sea are noticeably different for the surface temperature variable, while there is only a moderate difference for the precipitation variable. The smoothness parameter values from M3-4 are significantly different than those from M2-4 and M3-3 for the precipitation variable. Overall, the models are consistent in that precipitation process is smoother than the surface temperature process – except in the case of the surface temperature over the sea for the models M2-4, M3-3, and M3-4 – which agrees with the results in [13].

Table 4 shows the root mean squared errors (RMSE), mean absolute prediction errors (MAE), and continuous ranked probability scores (CRPS) for both variables over $L^{(i)}$, $i = 1, 2$, as well as a new set of prediction locations $P$. The locations

the data. Comparing M3-3 and M3-4, it can be seen that allowing the spatial range parameters for each $X_i^0$ to be different slightly improved the fit.
Table 3
Fitted smoothness parameter values (with asymptotic standard errors in parentheses) for the surface temperature and precipitation data under various models. M2-4, M3-3, and M3-4 yield different smoothness values over the land and the sea.

<table>
<thead>
<tr>
<th>IM</th>
<th>M1</th>
<th>M2-2</th>
<th>M2-4</th>
<th>M3-1</th>
<th>M3-2</th>
<th>M3-3</th>
<th>M3-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>Land</td>
<td>1.31 (0.036)</td>
<td>1.31 (0.188)</td>
<td>1.14 (0.120)</td>
<td>0.19 (0.078)</td>
<td>1.53 (0.100)</td>
<td>1.73 (0.099)</td>
</tr>
<tr>
<td></td>
<td>Sea</td>
<td>1.15 (0.046)</td>
<td>1.15 (0.036)</td>
<td>1.15 (0.188)</td>
<td>1.15 (0.120)</td>
<td>1.15 (0.078)</td>
<td>1.15 (0.099)</td>
</tr>
<tr>
<td>P</td>
<td>Land</td>
<td>3.00 (0.220)</td>
<td>2.99 (0.466)</td>
<td>1.81 (0.192)</td>
<td>1.37 (0.211)</td>
<td>1.80 (0.207)</td>
<td>2.01 (0.234)</td>
</tr>
<tr>
<td></td>
<td>Sea</td>
<td>2.91 (0.125)</td>
<td>2.91 (0.120)</td>
<td>2.91 (0.120)</td>
<td>2.91 (0.120)</td>
<td>2.91 (0.120)</td>
<td>2.91 (0.120)</td>
</tr>
</tbody>
</table>

Table 4
Root mean squared error (RMSE), mean absolute prediction error (MAE), and continuous ranked probability score (CRPS), for surface temperature and precipitation. The numbers in gray cells are the minimum for each row.

<table>
<thead>
<tr>
<th>IM</th>
<th>M1</th>
<th>M2-2</th>
<th>M2-3</th>
<th>M2-4</th>
<th>M3-1</th>
<th>M3-2</th>
<th>M3-3</th>
<th>M3-4</th>
<th>M4-1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>L</td>
<td>T</td>
<td>1.299</td>
<td>1.277</td>
<td>1.203</td>
<td>1.259</td>
<td>1.288</td>
<td>1.119</td>
<td>1.156</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>P</td>
<td>0.578</td>
<td>0.568</td>
<td>0.581</td>
<td>0.576</td>
<td>0.629</td>
<td>0.475</td>
<td>0.524</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td></td>
<td>1.493</td>
<td>1.497</td>
<td>1.576</td>
<td>1.438</td>
<td>1.446</td>
<td>1.320</td>
<td>1.354</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td></td>
<td>0.703</td>
<td>0.703</td>
<td>0.747</td>
<td>0.724</td>
<td>0.731</td>
<td>0.472</td>
<td>0.502</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.703</td>
<td>0.703</td>
<td>0.747</td>
<td>0.724</td>
<td>0.731</td>
<td>0.472</td>
<td>0.502</td>
</tr>
<tr>
<td>MAE</td>
<td>L</td>
<td>T</td>
<td>0.735</td>
<td>0.737</td>
<td>0.695</td>
<td>0.729</td>
<td>0.705</td>
<td>0.693</td>
<td>0.646</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>P</td>
<td>0.423</td>
<td>0.418</td>
<td>0.42</td>
<td>0.42</td>
<td>0.42</td>
<td>0.42</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td></td>
<td>0.739</td>
<td>0.74</td>
<td>0.751</td>
<td>0.734</td>
<td>0.673</td>
<td>0.692</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td></td>
<td>0.475</td>
<td>0.475</td>
<td>0.492</td>
<td>0.486</td>
<td>0.481</td>
<td>0.39</td>
<td>0.455</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.475</td>
<td>0.475</td>
<td>0.492</td>
<td>0.486</td>
<td>0.481</td>
<td>0.39</td>
<td>0.455</td>
</tr>
<tr>
<td>CRPS</td>
<td>L</td>
<td>T</td>
<td>0.323</td>
<td>0.323</td>
<td>0.332</td>
<td>0.364</td>
<td>0.313</td>
<td>0.442</td>
<td>0.417</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>P</td>
<td>0.284</td>
<td>0.284</td>
<td>0.062</td>
<td>0.161</td>
<td>0.318</td>
<td>0.12</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td></td>
<td>0.45</td>
<td>0.452</td>
<td>0.481</td>
<td>0.461</td>
<td>0.432</td>
<td>0.473</td>
<td>0.474</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td></td>
<td>0.467</td>
<td>0.467</td>
<td>0.423</td>
<td>0.426</td>
<td>0.482</td>
<td>0.458</td>
<td>0.491</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.467</td>
<td>0.467</td>
<td>0.423</td>
<td>0.426</td>
<td>0.482</td>
<td>0.458</td>
<td>0.491</td>
</tr>
</tbody>
</table>

in $\mathcal{P}$ are from a random sample of 100 locations disjoint from $\mathcal{L}$ and the $\mathcal{L}^{(i)}$'s. Unlike [11], which shows a better prediction performance for simpler models, we produced smaller prediction errors for the more elaborate covariance models. In particular, M3-4, the best model in terms of the AIC (as shown in Table 2), produces the smallest prediction errors in terms of the MAE, for the temperature variable. In fact, [2] note that having different spatial range parameters for each variable improves the prediction in some cases, as in the case of the temperature variable but not the precipitation variable. Note that the prediction errors over $\mathcal{L}^{(i)}$ are only slightly smaller than those over $\mathcal{P}$, which may be explained by a relatively weak cross-dependence of the two variables (see Section 5.2). It is interesting to note that M4-1, the model with the differential operators approach (but not with spatially varying smoothness), performs well, particularly for the precipitation variable.

5.2. Climate model error

Jun et al. [14] analyzed 19 error fields of climate model outputs for surface temperature variable to quantify their cross-dependence. They defined the difference between the climate model output and the corresponding observation as the error of the climate model and showed that, overall, many pairs of models have highly correlated errors and those developed by the same group have particularly highly correlated errors. They quantified such correlations using a kernel-based estimator based on the cross product of the two spatial error fields. In this paper, we build joint spatial models of pairs of error fields that respect the complex nonstationary structure of the variables. We use the two error fields from the climate models developed by the Geophysical Fluid Dynamics Laboratory in the US (GFDL-CM2.0) and the Hadley Center for Climate Prediction and Research in the UK (UKMO-HadCM3), and similarly to [14], we consider seasonal averages (December, January, and February) over 30 years (1970–1999) for the surface temperature variable. Although the two models are not developed by the same group, based on the result in [14] we expect a strong cross-correlation between the two model errors.

Similarly to Section 5.1, we randomly sampled 1000 locations for fitting the models and two sets of 100 random locations for each variable for prediction (we also ensured that only 10% of all the sampled locations were from the regions above 60°N and below 60°S). Fig. 4 shows the spatial map of the data. We can see a noticeable pattern depending on the latitude, and overall the data seems more variable over the land than over the sea. The overall level of co-located cross-correlation is quite high, but there is no clear pattern of a correlation depending on the latitude (not shown).

Table 5 shows the maximized log likelihood values, numbers of parameters, and AIC values for the models considered for this data set. The same description as for Table 1 applies here, except for a few details as explained below. Note that the data characteristics here are quite different from those in Section 5.1. In particular, the two climate model errors have similar spatial structures overall and the smoothness parameters do not have a significant dependence on whether the location is over the land or the sea. Furthermore, the cross-correlation between the two variables is quite high. Therefore, we do not

Table 5
Fitted smoothness parameter values (with asymptotic standard errors in parentheses) for the surface temperature and precipitation data under various models. M2-4, M3-3, and M3-4 yield different smoothness values over the land and the sea.
Fig. 4. Map of the two climate model errors. The figures on the left show the locations of $\mathcal{L}$ and those on the right show the locations of $\mathcal{L}^{(i)}$ ($i = 1$ for GFDL-CM2.0 and $i = 2$ for UKMO-HadCM3).

Table 5

Maximum loglikelihood values, numbers of parameters, and AIC values for the covariance models fitted to the two climate model errors for the temperature variable. The maximum loglikelihood value, numbers of parameters, and AIC value, for the independent, isotropic Matérn model are $-2097.91$, $5$, and $-4205.82$, respectively. $\Delta \text{loglik}$ gives the increase in the loglik values and $\Delta \text{AIC}$ gives the decrease in the AIC values, both compared to IM.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>M1</th>
<th>M2-2</th>
<th>M3-2</th>
<th>M4-1</th>
<th>M4-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \text{loglik}$</td>
<td>206.66</td>
<td>687.23</td>
<td>505.76</td>
<td>668.2</td>
<td>719.323</td>
</tr>
<tr>
<td>$\Delta \text{AIC}$</td>
<td>411.32</td>
<td>1364.46</td>
<td>983.52</td>
<td>1298.4</td>
<td>1400.65</td>
</tr>
</tbody>
</table>
Table 6
Fitted smoothness parameter values (with their asymptotic standard errors in parentheses) for the two climate model errors. For M4-2, the values in the table are for $Y_1$ and $Y_2$ (the form of M4-2 is given in Table 1). The smoothness estimate for $X_i$ in M4-2 is 1.47 (0.405).

<table>
<thead>
<tr>
<th></th>
<th>IM</th>
<th>M1</th>
<th>M2-2</th>
<th>M3-2</th>
<th>M4-1</th>
<th>M4-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFDL-CM2.0</td>
<td>0.63(0.002)</td>
<td>0.63(0.039)</td>
<td>0.63(0.042)</td>
<td>1.34(0.032)</td>
<td>0.50(0.052)</td>
<td></td>
</tr>
<tr>
<td>UKMO-HadCM3</td>
<td>0.28(0.001)</td>
<td>0.28(0.023)</td>
<td>0.35(0.025)</td>
<td>1.13(0.021)</td>
<td>0.29(0.028)</td>
<td></td>
</tr>
</tbody>
</table>

Table 7
Root mean squared errors (RMSE), mean absolute prediction errors (MAE), and continuous ranked probability scores (CRPS), for the climate model errors. Numbers in gray cells are the minimum for each row.

<table>
<thead>
<tr>
<th></th>
<th>IM</th>
<th>M1</th>
<th>M2-2</th>
<th>M3-2</th>
<th>M4-1</th>
<th>M4-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE $L^{(1)}$</td>
<td>GFDL-CM2.0</td>
<td>0.94</td>
<td>0.84</td>
<td>0.839</td>
<td>0.903</td>
<td>0.845</td>
</tr>
<tr>
<td></td>
<td>UKMO-HadCM3</td>
<td>3.498</td>
<td>2.616</td>
<td>2.624</td>
<td>3.006</td>
<td>2.963</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>GFDL-CM2.0</td>
<td>7.396</td>
<td>7.367</td>
<td>7.361</td>
<td>7.205</td>
</tr>
<tr>
<td>MAE $L^{(1)}$</td>
<td>GFDL-CM2.0</td>
<td>0.63</td>
<td>0.591</td>
<td>0.584</td>
<td>0.605</td>
<td>0.607</td>
</tr>
<tr>
<td></td>
<td>UKMO-HadCM3</td>
<td>1.075</td>
<td>0.947</td>
<td>0.931</td>
<td>0.969</td>
<td>1.035</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>GFDL-CM2.0</td>
<td>2.097</td>
<td>2.095</td>
<td>2.094</td>
<td>2.054</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UKMO-HadCM3</td>
<td>2.228</td>
<td>2.205</td>
<td>2.224</td>
<td>2.197</td>
</tr>
<tr>
<td>CRPS $L^{(1)}$</td>
<td>GFDL-CM2.0</td>
<td>2.34</td>
<td>2.34</td>
<td>2.349</td>
<td>2.393</td>
<td>2.358</td>
</tr>
<tr>
<td></td>
<td>UKMO-HadCM3</td>
<td>1.112</td>
<td>1.112</td>
<td>1.021</td>
<td>2.109</td>
<td>0.969</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>GFDL-CM2.0</td>
<td>1.402</td>
<td>1.405</td>
<td>1.406</td>
<td>1.359</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UKMO-HadCM3</td>
<td>0.86</td>
<td>0.854</td>
<td>0.853</td>
<td>1.097</td>
</tr>
</tbody>
</table>

and M2-2, again unlike the example in Section 5.1; note that the effective smoothness parameters for M3-2 should be the values in Table 6 minus 1 since the differential operators effectively reduce the smoothness values by 1.

Table 7 gives the root mean squared errors (RMSE), mean absolute prediction errors (MAE), and continuous ranked probability scores (CRPS). Similarly to the data set in Section 5.1, the simpler cross-covariance models provided poorer predictions than the more elaborate models, except for the RMSE of the error of UKMO-HadCM3 for M1 and the CRPS of the error of GFDL-CM2.0 for IM and M1. It is interesting that models M3-2 and M4-2 do particularly well for GFDL-CM2.0, while model M4-1 does well for UKMO-HadCM3. Since the cross-dependence of the two variables is quite strong, the prediction errors over the $L^{(1)}$‘s are much smaller than those over $P$.

5.3. Numerical issues

To reduce the computational burden for both examples in Sections 5.1 and 5.2, we randomly sampled 1000 locations for the estimation. However, the estimated values of some of the covariance parameters – in particular, the smoothness parameters – may be affected by the sampling locations. That is, when we have clusters of locations rather than random samples of locations over the entire domain, some of the smoothness parameter estimates may change significantly. As a sensitivity check, we fitted a few previously considered covariance models using data over different sampling locations. For the first example, we fitted data from the locations over North America (a total of 1152 locations), and for the second example, we used essentially all the data points. In terms of the comparisons regarding maximum loglikelihood and AIC values, the results were similar to the previous results. For example, the maximum loglikelihood values of M1, M2-2, and M3-4 for the first data set are 1780.42, 2121.967, and 2351.89, respectively, while those of M1, M2-2, and M4-2 for the second data set are $-3216.24$, $-2051.93$, and $-2000.36$, respectively.

For the first example, the smoothness parameter estimates (with their asymptotic standard errors) are 0.39 (0.017) and 1.12 (0.023), for temperature and precipitation, respectively, in M1. In M3-4, for temperature, the estimates for the smoothness parameters are 1.442 (0.021) and 1.446 (0.031) over land and sea, respectively. For precipitation, the estimates are 1.517 (0.020) and 1.337 (0.038) over land and sea, respectively. For this particular data set, the smoothness estimates changed noticeably compared to the result in Table 3. Note that the asymptotic standard errors are reduced compared to Table 3, which was expected given that there are denser locations. For the second example, the estimates are 0.36 (0.015) and 0.15 (0.009) for GFDL-CM2.0 and UKMO-HadCM3, respectively, in M1. In M2-2, the estimates are 0.61 (0.026) and 0.34 (0.021), respectively. Similar to the first example, the estimated asymptotic standard errors were significantly reduced significantly. Furthermore, unlike the first example, the estimates for the smoothness parameters are consistent with the values in Table 6.
All of the maximum likelihood estimation results presented in this paper were done numerically using the \texttt{nlm} and \texttt{optim} functions in R. We made certain that both optimization functions obtain a stable numerical minimum (for the negative loglikelihood function) in all cases, and we additionally checked that the Hessian matrices are invertible, resulting in the asymptotic standard errors for each parameter estimate. None of the results presented in 5.1 and 5.2 exhibited numerical instability in this respect. Using Intel Xeon CPU 2.90 GHz, calculation of the full loglikelihood was performed in a reasonable amount of time. For example, in the first example, the loglikelihood calculation for model M3-4 with 22 parameters took 20 s and the numerical optimization using \texttt{nlm} and \texttt{optim} took at most 1–2 days.

In the sensitivity check, however, there were numerical issues in the optimization for M2-2 in the first example and for M4-2 in the second example. For M2-2 in the first example, one particular parameter, the linear term of latitude in $\sigma_i$ (for temperature) was approaching the boundary, and the \texttt{nlm} routine returned a code claiming the function was too non-linear. Interestingly, \texttt{optim} claimed proper convergence even though the asymptotic variance for the particular parameter approaching the boundary turned out to be negative (a sign of numerical instability). On the other hand, for M4-2 in the second example, although the optimization routine converged (and the \texttt{nlm} function did not return any message noting convergence problems), the smoothness parameter for the $X_i$s was quite small and the corresponding asymptotic variance from Hessian was negative. Although the optimization routine succeeded, this is again a sign of numerical instability.

6. Discussion

We presented several nonstationary Matérn-based covariance models for multivariate processes on a sphere. In particular, we demonstrated the effectiveness of spatially varying smoothness parameters in the multivariate Matérn model using the data in [11]. We showed how the nonstationary Matérn model with spatially varying smoothness parameters can be coupled with the differential operators approach, to achieve more flexible parametric cross-covariance models. Regarding the differential operators approach, we proposed approaches that respect the curvature of the earth and some models that are mean-square continuous everywhere on the sphere. These models were tested in terms of model fitting and prediction through two climate data examples.

When the smoothness parameters of the $Z_i$s vary over space, they obviously have a dependence on latitude and longitude. Therefore, in (7), the covariance function of the $Z_i$s needs to involve derivatives of the smoothness parameters with respect to latitude and longitude. Page 254 of [22] provides the expression for the derivative of a modified Bessel function with respect to its order (i.e., smoothness parameter), but the expression is complex and involves infinite summations. Covariance models such as M2-3 have smoothness parameters that are explicit functions of the latitude. Thus, when we couple the differential operators approach with M2-3, we may have to consider the derivative of smoothness parameters with respect to latitude. However, in this paper we only considered M2-4 coupled with the differential operators approach, and in this case the smoothness parameters are constant over land and sea (with different values in the two cases). Therefore, differentiating the smoothness parameter with respect to latitude may only make a difference near coastlines. Furthermore, it is not clear how to explicitly express the smoothness parameters as functions of latitude or longitude in this case. Therefore, in calculating the covariance of the coupled models M3-3 and M3-4, we have proceeded as if the smoothness parameters are fixed over space and have excluded partial derivatives of the smoothness parameters with respect to latitude and longitude. It would be interesting to determine how much difference this makes through either analytical calculation or numerical approximation of the derivatives, and we leave this for future work.

The idea of coupling the differential operators approach with isotropic or nonstationary cross-covariance models can be applied to any cross-covariance functions other than multivariate Matérn model. Due to the computational burden, we randomly sampled a manageable size of spatial locations to fit the joint model through maximum likelihood estimation. To fit the full data, covariance approximation or likelihood approximation methods [31,3,8,6,27] in the recent literature may be employed.

As discussed in Section 1, an issue with modeling processes on a sphere with the Matérn class, either isotropic or nonstationary, is that the smoothness parameters cannot exceed 0.5. In this paper, we used chordal distance to avoid this issue, although it may not be a natural metric on the sphere. To the best of the author’s knowledge, there has not been much work in the literature on valid parametric covariance functions, like the Matérn class, that do not involve infinite summations with great circle distances, other than the work of Gneiting [10]. This is another research direction that the author is currently pursuing.

Acknowledgments

Mikyoung Jun’s research was supported by NSF grant DMS-1208421. This publication is based in part on work supported by Award No. KUS-C1-016-04, made by King Abdullah University of Science and Technology (KAUST). The author acknowledges the modeling groups for making their simulations available for analysis, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving the CMIP3 model output, and the World Climate Research Programme (WCRP)’s Working Group on Coupled Modelling (WGCM) for organizing the model data analysis activity. The WCRP CMIP3 multi-model data set is supported by the Office of Science, U.S. Department of Energy. The author thanks the Editor, the Associate Editor, and two anonymous reviewers whose comments helped to improve the paper significantly.
Appendix. Proof of Proposition 1

Similarly to [13], if a bivariate process on a sphere, \( Z = (Z_1, Z_2) \), has the covariance structure in (1), we can write

\[
\text{cov}(Z_1(l_1, l_1), Z_2(l_2, l_2)) = \Gamma_1 \mathcal{M}_{1-2}(h^{1/2}) + \Gamma_2 \mathcal{M}_{1-1}(h^{1/2}),
\]

where \( h = h(l_1, l_2, l_1 - l_2) = (D/\xi)^2 \), \( D \) is the chordal distance, \( \xi \) is the spatial range parameter for \( Z \) in (1),

\[
\Gamma_1 = \{A_{11}(l_1)A_{21}(l_2)h_1h_2 - B_{11}(l_1)B_{21}(l_2)h_1^2 - A_{11}(l_1)B_{21}(l_2)h_1h_3 + B_{11}(l_1)A_{21}(l_2)h_2h_3\}/4,
\]

\[
\Gamma_2 = -\{A_{11}(l_1)A_{21}(l_2)h_1h_2 - B_{11}(l_1)B_{21}(l_2)h_1^2 - A_{11}(l_1)B_{21}(l_2)h_1h_3 + B_{11}(l_1)A_{21}(l_2)h_2h_3\}/2,
\]

and \( h_0 = h(l_1, l_2, l_1 - l_2) = \partial h/\partial x_p \), \( h_0 = h_p(l_1, l_2, l_1 - l_2) = \partial^2 h/\partial x_p \partial x_q \) (\( x_1 = l_1, x_2 = l_2, x_3 = l_1 - l_2 \), see Appendix A of [15] for explicit expressions for \( h_p \) and \( h_0 \)), and \( \mathcal{M}_p(x) = x^p \mathcal{X}_p(x) \). Then, when \( l_1 = l_2 = L \) and \( l_1 = l_2 = l \), we have \( \Gamma_1 = 0 \) and \( \Gamma_2 = R^2 \{A_{11}(L)A_{21}(L) + B_{11}(L)B_{21}(L)(\cos^2 L)\} \). The rest of the result is immediate.

References